LCDR Dylan Schmorrow
Information Processing Technology Office (IPTO)
Small Business Innovation Research (SBIR) Speech

Ladies and Gentlemen, Good morning. My name is LCDR Dylan Schmorrow and I am a Program manager in the Information Processing Technology Office.

One of the most rewarding tasks of a DARPA program manger is participation in the advancement of small businesses that ordinarily would not have the internal funds to develop technology. Small Business Innovation Research, or SBIR, targets the entrepreneurial sector where the risk and expense of conducting serious research and development efforts are often beyond the means of many small businesses. By reserving selected federal research and development funds for small business, SBIR funds the critical startup and development stages and encourages the commercialization of the technology, product, or service.

The DARPA-sponsored project at Insightful Corporation, "An Inverse Inference Engine for High Precision Web Search," has lead to the development of a new product known as InFact, an advanced search engine based on linguistic normalization. The research team has developed sophisticated technology for modeling the morphology, semantics, and syntax of languages. It captures a larger amount of information than other search engines and can "read" documents while retaining the information content of every sentence.

The first generation of search engines retrieved information by keywords. The next generation attempted to retrieve information by concepts. The third generation of search engines will retrieve facts. In the past, search engines scanned pages for keywords and patterns, but InFact understands meanings and duplicates the process of lexical comprehension in humans.

This technology uses language models that examine morphology for the analysis of word form, syntax for the analysis of keyword relationships, and semantics for the analysis of overall linguistic meaning. Morphological and semantic analyses are not new to search and retrieval but it is the inclusion of the syntactic model that makes this product unique.

I will now share a short video describing the company and their DARPA-funded work.

<<PLAY VIDEO 4min 42 sec>>

As you saw, technological capabilities include:

- Natural language question answering,

- Summarization of facts across multiple sources,

- Transformation of raw information to ideas,

- Understanding of text documents as well as images, charts, and graphs,

- Incremental indexing of information,

- Automatic categorization identifying themes within information, and

- Multilingual speech recognition capabilities.

InFact uses a process of inductive reasoning to learn word meanings from their context and can discriminate between concepts that are like or unlike in seconds. The patented algorithm, Latent Semantic Regression, is

a computational breakthrough that increases searching power by understanding content rather than looking for keywords or phrases.

In addition to the semantic network, rule-based and stochastic parsers are employed to recognize grammatical structures. The governing parts of speech and their modifiers can therefore be identified, better capturing the meaning for each sentence. This information is mapped into patented smart data structures that are modeled after the notion of transformational grammars. This allows sentence similarity searches, cross-document text summarization, and advanced searches where keywords can be constrained to syntactic roles.

The development of this high precision web search tool is a good example of DARPA research: work that results in new and innovative solutions to widely occurring problems. InFact is representative of an entirely new approach to the next generation of search engines and would not have been possible without critical research and development funds from DARPA. With the ever-increasing amount of data available, it is becoming more important every day to have the ability to sort through that data and extract what is important. This work has helped advance the state of the art in this technology area and will help us better navigate the growing sea of information around us.

Thank you.